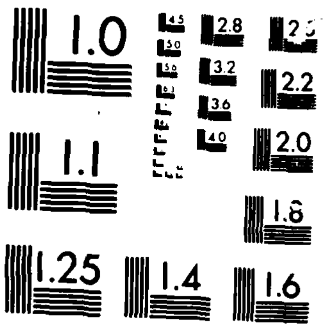
AD-A166 148 STOCHASTIC SYSTEMS WITH SMALL NOISE ANALYSIS AND SIMULATION: A PHASE LOCK..(U) BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMICAL SYSTE.. P DUPUIS ET AL. 1/1

UNCLASSIFIED AUG 85 LCDS-85-20 AFOSR-TR-86-0033 F/G 9/4 NL



MICROCOPY RESOLUTION TEST CHART

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(2)

AD-A166 148

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 86-0033 ~~LCDS #85-20~~	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Stochastic Systems With Small Noise, Analysis and Simulation: A Phase Locked Loop Example		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) P. Dupuis and H. Kushner		8. CONTRACT OR GRANT NUMBER(s) [redacted]-AFOSR 81-0116
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lefschetz Center for Dynamical Systems Division of Applied Mathematics Brown University, Providence, RI 02912		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304 A1
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR/NM Bolling Air Force Base Washington, DC 20332		12. REPORT DATE August 1985
		13. NUMBER OF PAGES 40
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Included

86 4 1 036

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract

Systems with wide bandwidth noise inputs are a common occurrence in stochastic control and communication theory and elsewhere; e.g., tracking or synchronization systems such as phase locked loops (PLL). One is often interested in calculating such quantities as the probability of escape from a desired 'error' set, in some time interval, or the mean time for such escape. Diffusion approximations (the system obtained in the limit bandwidth $\to \infty$) are often used for this, being easier to analyze. When the noise effects in the physical system are small, one is tempted to do an asymptotic analysis (noise intensity $\to 0$) on the diffusion approximation, and use this for the desired estimates on the original system. Such a procedure does not work in general: the double limit bandwidth $\to \infty$, intensity $\to 0$ is not always justified. Under quite broad conditions on the noise processes, it is justified for the systems studied here. We study a particular form of the PLL owing to the great practical importance of the system and because it provides a useful vehicle for understanding the extent of validity of the asymptotic methods for such systems. The basic analytical techniques are from the theory of large deviations. One seeks information on the escape probabilities, mean times, and on the most likely exit paths and exit locations. Also, we seek information on the interactions between the signals to be tracked and the noise which are most likely to lead to exit. The large deviations technique is eminently suited to this job.

Simulations are taken in order to understand the range of validity of the asymptotic method. Agreement between the predictions and sample estimates is good over noise intensity levels which seem to be ever larger than

those typically occurring in practice. Since the events whose probability is of interest have small probability, an importance sampling scheme is used to 'quicken' the simulations. The scheme works well. The required measure transformations are suggested by the theory of large deviations, and are obtained by solving the associated variational problem. The technique seems to be very appropriate as both an analysis and design tool for such systems.

Key Words: large deviations, approximations of systems with wide bandwidth noise, importance sampling, quick simulation, small noise diffusion models, asymptotic methods, phase locked loops, synchronization systems.

2

STOCHASTIC SYSTEMS WITH SMALL NOISE, ANALYSIS AND SIMULATION; A PHASE LOCKED LOOP EXAMPLE

by

P. Dupuis and H. Kushner

August 1985 LCDS #85-20

STOCHASTIC SYSTEMS WITH SMALL NOISE, ANALYSIS AND SIMULATION; A PHASE LOCKED LOOP EXAMPLE

P. Dupuis*
H. Kushner**

Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

August 1985

* This reserach was supported in part by the Army Research Office under Contract No. DAAG 29-84-K-0082, the Office of Naval Research under Contract No. N00014-83-K-0542, and the National Science Foundation under Grant No. DMS 85-11470.

** This research was supported in part by the National Science Foundation under Grant No. ECS 82-11476 and the Air Force Office of Scientific Research under Contract No. AF-AFOSR 81-0116.

Abstract

Systems with wide bandwidth noise inputs are a common occurrence in stochastic control and communication theory and elsewhere; e.g., tracking or synchronization systems such as phase locked loops (PLL). One is often interested in calculating such quantities as the probability of escape from a desired 'error' set, in some time interval, or the mean time for such escape. Diffusion approximations (the system obtained in the limit bandwidth $\to \infty$) are often used for this, being easier to analyze. When the noise effects in the physical system are small, one is tempted to do an asymptotic analysis (noise intensity $\to 0$) on the diffusion approximation, and use this for the desired estimates on the original system. Such a procedure does not work in general: the double limit bandwidth $\to \infty$, intensity $\to 0$ is not always justified. Under quite broad conditions on the noise processes, it is justified for the systems studied here. We study a particular form of the PLL owing to the great practical importance of the system and because it provides a useful vehicle for understanding the extent of validity of the asymptotic methods for such systems. The basic analytical techniques are from the theory of large deviations. One seeks information on the escape probabilities, mean times, and on the most likely exit paths and exit locations. Also, we seek information on the interactions between the signals to be tracked and the noise which are most likely to lead to exit. The large deviations technique is eminently suited to this job.

Simulations are taken in order to understand the range of validity of the asymptotic method. Agreement between the predictions and sample estimates is good over noise intensity levels which seem to be ever larger than

those typically occurring in practice. Since the events whose probability is of interest have small probability, an importance sampling scheme is used to 'quicken' the simulations. The scheme works well. The required measure transformations are suggested by the theory of large deviations, and are obtained by solving the associated variational problem. The technique seems to be very appropriate as both an analysis and design tool for such systems.

Key Words: large deviations, approximations of systems with wide bandwidth noise, importance sampling, quick simulation, small noise diffusion models, asymptotic methods, phase locked loops, synchronization systems.

I. Introduction

Let $\xi^\lambda(\cdot)$ denote a wide bandwidth (BW) process (BW $\to \infty$ as $\lambda \to 0$), $\{\xi_n^\lambda, 0 \le n < \infty\}$ a sequence of correlated random variables for each $\lambda > 0$, and $w(\cdot)$ a standard Wiener process. Systems of the types (1.1) to (1.3) are frequently used models in stochastic control and communication theory and elsewhere

(1.1) $\quad \dot{x}^\lambda = F_\lambda(x^\lambda, \xi^\lambda)$

(1.2) $\quad X_{n+1}^\lambda = X_n^\lambda + \lambda F_\lambda(X_n^\lambda, \xi_n^\lambda)$

(1.3) $\quad dx = f(x)dt + \sigma(x)dw \quad .$

The function F_λ in (1.1) is such that for small λ, the process $x^\lambda(\cdot)$ is close (say, in the sense of weak convergence) to the diffusion $x(\cdot)$. Similarly for the F_λ in (1.2).

Typically, (1.3) is used as an approximation to the 'physical' models (1.1) or (1.2). Signal tracking and synchronization systems (with noise corrupted observations) represent one very important class of applications in communication theory; see, for example, the enormous literature on phase locked loops (PLL), [1] to [4]. For such systems, one or more of the components of the state x in (1.1) might represent the error in tracking a signal, and the others might represent internal states of the system (the filter, etc.). Typically, one is interested in estimates of quantities such as

(1.4) $\quad P_x\{x^\lambda(t) \notin G, \text{ some } t \le T\}$

(1.5) $\quad E_x \tau_G^\lambda$, where $\tau_G^\lambda = \inf\{t: x^\lambda(t) \notin G\} \quad ,$

for suitable sets G and times T. Estimates of such quantities are generally quite hard to get for any of the models (1.1) to (1.3), but certainly much harder (if not impossible) for the wide BW driven system (1.1) (or (1.2)), than for (1.3). Owing to this, one is often forced to use some sort of asymptotic method.

Since (1.3) is much easier to work with than (1.1) or (1.2) one is strongly tempted to find a diffusion approximation to $x^\lambda(\cdot)$ or to $\{X_i^\lambda\}$. If the noise effects are not 'small,' then the diffusion approximation method can give good results [19], [20]. But even for this case, it might be hard to solve for the analogs of (1.4), (1.5):

(1.4') $\quad P_x\{x(t) \notin G, \text{ some } t \le T\}$

(1.5') $\quad E_x\tau_G\,,\quad \tau_G = \inf\{t: x(t) \notin G\}$.

Often, one estimates quantities such as (1.4'), (1.5') by assuming that the noise effects are small; e.g., by using the model (1.6), and estimating (1.7), (1.8) via some other asymptotic method.

(1.6) $\quad dx^\varepsilon = f(x^\varepsilon)dt + \varepsilon\sigma(x^\varepsilon)dw$

(1.7) $\quad P\{x^\varepsilon(t) \notin G, \text{ some } t \le T\}$

(1.8) $\quad E_x\tau_G^\varepsilon\,,\quad \tau_G^\varepsilon = \inf\{t: x^\varepsilon(t) \notin G\}$.

We have mentioned two very different types of approximations. First, letting $BW \to \infty$ to approximate (1.1) by (1.3), then letting the noise intensity go to zero (replacing σ by $\varepsilon\sigma$) to get an asymptotic estimate for (1.7), (1.8). This procedure is rather risky in that the estimates of (1.7), (1.8) might be very poor estimates of (1.4), (1.5), even when the

'intensity' of the noise effects in (1.1) is small. See [5] for some examples. Basically, when ε or the noise effects in (1.1) are small, the events of interests are rare (e.g., the event in (1.4) or (1.7)) or the times (1.5), (1.8) very large. The escape phenomenon depends on rarely occurring large bursts of noise, and the usual 'central limit theorem like' arguments which are used to model (1.1) by (1.3) do not model these well.

In order to use an asymptotic expansion based on (1.6) for estimates of functionals of (1.1), one must show that it yields the same results as one would get from (1.1) if the BW went to ∞ and the noise intensity went to zero simultaneously (with an analogous result for (1.2)), <u>not first letting</u> BW $\to \infty$, <u>then</u> intensity $\to 0$. For a large class of noise processes, such results are in [6]. These results seem to be crucial if one is to employ asymptotic methods for (1.6) to get properties of (1.1).

One interesting asymptotic method based on (often formal) expansions and boundary layer matchings is discussed in [7]-[8]. A rigorous approach to such a scheme involves strong regularity and ellipticity conditions on the differential generator of (1.3). An alternative approach is via the theory of large deviations, where the estimates (1.4), (1.5) or (1.7), (1.8), involve the solution to a variational problem [6], [9], [10], [11]. A related small noise expansion is in [12].

Once the small noise (small ε) expansion is obtained, one must ask how good it is for values of ε which one might expect in applications. To get such information, there seems (at the moment) little alternative to simulation. The questions of model approximation and validation by simulation will be addressed here for one particular PLL. Our PLL model is chosen both because it is important in applications and because it provides a good illustration of the general method.

In Section 2, we formulate the particular PLL problem of interest here. The diffusion approximation form of this system is <u>not</u> of the form (1.3) unless various ('double frequency') terms are dropped, the system reduced to 'baseband' form, and the wide bandwidth 'small intensity' observation noise replaced by small intensity white Gaussian noise. This procedure is, in fact, valid for our case, under quite broad conditions. Reference [13] deals with a related problem, and below we comment further on the relations between that work and ours. The basic scheme of the theory of large deviations and the simplifications of our model is discussed in Section 3.

Generally, with small noise effects, the 'escape' probabilities of the forms (1.4) (or (1.7)) are very small; conversely (1.5) (or (1.8)) would be quite large. In order to get good simulation results with a reasonable amount of computational effort, a form of importance sampling is used. The idea is to change the measure on the original probability space in a way which increases the probability of the 'rare' escape event on the time interval $[0,T]$. Then, we get the required estimate by an inverse transformation. A natural way in which to do this is suggested by the mathematics of the theory of large deviations, and (as will be seen) works very well. The idea was first formally used in [14], and is discussed in Section 4. Simulation results for the PLL (with a constant phase process) appear in Section 5. It is remarkable that the estimates given by the asymptotic theory are quite good even for rather large values of ϵ.

Let $\theta(\cdot)$ and $\hat{\theta}^{\varepsilon}(\cdot)$ (for model (1.6), or $\hat{\theta}^{\lambda}(\cdot)$ if the model is (1.1)) denote the phase to be tracked and its estimate, as given by the PLL. The error $\theta(\cdot) - \hat{\theta}^{\varepsilon}(\cdot)$ (or $\theta(\cdot) - \hat{\theta}^{\lambda}(\cdot)$) will be a component of the state x, and we use sets G of roughly the form $G = G_0 = \{x: |\theta-\hat{\theta}| < \pi\}$.

This set is standard in the study of the PLL, since when the error reaches π- the system can 'lose track.' An advantage of the large deviations approach is that one can (rigorously) get information on the most likely locations of points of escape from G, on the paths in whose (small) neighborhoods escape is most likely, and on the most likely magnitudes of the noise effects when escape occurs. When the signal is time varying, the scheme gives the most likely interactions between the signal and noise which lead to escape.

In Sections 6 and 7, we discuss the results when the phase varies as it would in a digital pulse phase modulation system. The model and the simulations are discussed and one can clearly see the interactions between the anticipated changes in the signal and the noise processes which are most likely to lead to escape.

An interesting study of a related problem is in [13]. The PLL in [13] is somewhat more complicated than ours. They start with a 'baseband' system of 4th order and reduce it to a 2nd order system. The input signal has wide BW of $O(K)$, where k is also a system gain, and the observation noise intensity ($\sqrt{\varepsilon}$ is used there rather than ε) is small. Since the system is not of the form (1.6), a scale change in both time and state is made (in an ε, k-dependent way) in order to get a diffusion model whose driving Wiener process is multiplied by a small parameter. Let L^{ε} denote the differential generator of the rescaled system--which we can suppose has the form (1.6). Let $V^{\varepsilon}(x) = E_x\tau^{\varepsilon}$. Then, formally, $L^{\varepsilon}V^{\varepsilon}(x) + 1 = 0$, $x \in G$, $V^{\varepsilon}(x) = 0$, $x \in \partial G$. Write $V^{\varepsilon}(x) = \exp\left[\frac{H}{\varepsilon} + C(x) + \text{higher order terms}\right]$. Via a formal expansion and boundary layer matching method, an interesting (but heuristic) approximation for $C(x)$ and H are obtained. Their simulations were in good agreement with the theoretical predictions over the

range of ε used. Although the systems are hard to compare, it seems that our range of ε (or at least our range of noise effects) is somewhat larger, if one uses (e.g.) estimates of (1.7), (1.8) as a measure of noise effects. The results are hard to compare directly with ours, since the system types are quite different, and because of the way ε and k are buried in the time and state scaling in [13].

Our method would not get $C(x)$, but the importance of this term, in general, is not clear. If ε is large, then the (small ε) approximation to this term might be unreliable. If ε is small, the overwhelming effects are due to $\exp H/\varepsilon^2$.

2. The PLL: The Model and Approximation

The basic physical system model is in Fig. 2.1. The VCO (voltage controlled oscillator) is a device whose output deviates from a reference frequency according to the input voltage. In order to get an asymptotic analysis from a reasonably practical perspective we proceed as follows, using the fact that for many practical systems, the carrier frequency is high and the BW narrow relative to the carrier but high in an absolute sense. We use a (high) carrier frequency ω^γ and an observation noise process whose BW is of the order of $1/\gamma$, but is small relative to ω^γ. Let η_γ be such that $\eta_\gamma/\gamma \to 0$ as $\gamma \to 0$, and let $\omega^\gamma = \omega_0/\eta^\gamma$ for some ω_0.

Let $\xi_i^\gamma(\cdot)$, $i = 1.2$, be mutually independent zero mean second order stationary processes. The BW of $\xi_i^\gamma(\cdot)$ will be of $0(1/\gamma)$. Let $\rho = (\varepsilon,\gamma)$. Following a common practice in communication theory, we model the observation noise in the 'passband' form $n^\rho(t) = \varepsilon u^\gamma(t)$, where

$$(2.1) \qquad u^\gamma(t) = \xi_1^\gamma(t) \sin \omega^\gamma t + \xi_2^\gamma(t) \cos \omega^\gamma t \quad .$$

The ε indexes the intensity of the noise and γ the BW. The following two models for $\xi_i^\gamma(\cdot)$ can be used and cover many cases in practice. For case 1, let $\xi_i^\gamma(t) = \xi_i(t/\gamma)/\sqrt{\gamma}$, where $\xi_i(\cdot)$ is a component of a stationary Gauss-Markov process with an integrable correlation function. Write $w_i^\gamma(t) = \int_0^t \xi_i^\gamma(s)\, ds$. Then the pair $\left(w_1^\gamma(\cdot), w_2^\gamma(\cdot)\right)$ converges weakly to a pair of mutually independent Wiener processes, each of which has covariance $\int_{-\infty}^{\infty} E\xi(s)\xi(0)ds = \sigma^2$. This is one standard way of modelling wide BW noise. We prefer to separate the parameters for intensity and BW, although they can be combined. Similarly for the case below.

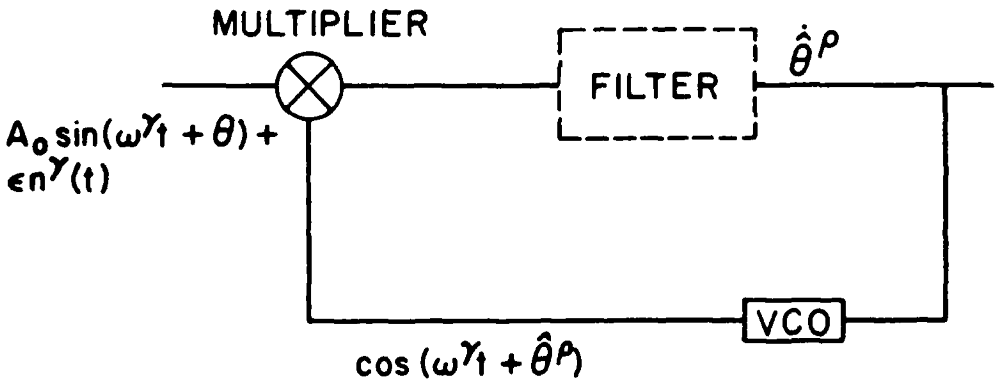


FIGURE 2.1 A PHASE LOCKED LOOP

Much physical noise has an 'impulsive' origin. The 'raw' impulses are either filtered by a circuit or system, or else the impulses are actually pulses whose intensity decreases rapidly. To model this, we use the second case where $\xi_i^\gamma(t) = \bar{\xi}_i^\gamma(t/\gamma)/\gamma$. where $\bar{\xi}_i^\gamma(\cdot)$ is a filtered impulsive noise process, the jump rate and moments depending on γ (increasing and decreasing, respectively, as $\gamma \to 0$). The specific model is Case II of [6]. The pair $\left(w_1^\gamma(\cdot), w_2^\gamma(\cdot)\right)$ again converges (under appropriate conditions on the jump rate and moments) to a pair of independent Wiener processes with covariance σ^2. See [6] for more detail.

The input to our PLL (signal plus noise) is

$$\tilde{u}^\rho(t) = A_0 \sin(\omega^\gamma t + \theta) + \varepsilon u^\gamma(t)$$

and the system equations are (see Fig. 1)

$$\text{(2.2)} \qquad \begin{aligned} \dot{v}^\rho &= A v^\rho + B \cos(\omega^\gamma t + \hat{\theta}^\rho)\tilde{u}^\rho(t) \\ \dot{\hat{\theta}}^\rho &= H v^\rho \quad , \end{aligned}$$

where A is a stable matrix. Owing to the complicated form of the expression (2.3), (2.2) is not suitable for analysis or computation.

$$\text{(2.3)} \qquad \begin{aligned} &\cos(\omega^\gamma t + \hat{\theta}^\rho)\tilde{u}^\rho(t) = \\ &\quad \frac{A_0}{2} \sin(\theta - \hat{\theta}^\rho) + \frac{A_0}{2} \sin(\theta + \hat{\theta}^\rho + 2\omega^\gamma t) \\ &\quad + \frac{\varepsilon}{2} [\xi_1^\gamma(t) \cos \hat{\theta}^\rho - \xi_2^\gamma(t) \sin \hat{\theta}^\rho] \\ &\quad + \frac{\varepsilon}{2} [\xi_1^\gamma(t) \cos(2\omega^\gamma t + \hat{\theta}^\rho) + \xi_2^\gamma(t) \sin(2\omega^\gamma t + \hat{\theta}^\rho)] \quad . \end{aligned}$$

If we were justified in dropping the parts of (2.3) containing ω^γ, then we could replace (2.2) by the system

$$\dot{\tilde{v}}^\rho = A\tilde{v}^\rho + B\left\{\frac{A_0}{2}\sin(\theta-\tilde{\theta}^\rho)\right.$$

$$\left. + \frac{\varepsilon}{2}\left[\xi_1^\gamma(t)\cos\tilde{\theta}^\rho - \xi_2^\gamma(t)\sin\tilde{\theta}^\rho\right]\right\} , \tag{2.4}$$

$$\dot{\tilde{\theta}}^\rho = H\tilde{v}^\rho .$$

If, furthermore, we were justified in replacing the $\xi_1^\gamma(\cdot)$ in (2.4) by white Gaussian noise of intensity σ, we could use the model (2.5) for the 'asymptotic' calculations.

$$dv^\varepsilon = Av^\varepsilon dt + B\frac{A_0}{2}\sin(\theta-\hat{\theta}^\varepsilon)dt \tag{2.5}$$

$$+ B\frac{\varepsilon}{2}\sigma dw$$

$$\dot{\hat{\theta}}^\varepsilon = Hv^\varepsilon , \quad (\theta-\hat{\theta}^\varepsilon)^{\bullet} = -Hv^\varepsilon \quad \text{if} \quad \theta(t) = \text{constant}.$$

It turns out that one <u>can</u>, in fact, make all of these approximations in the sense that the asymptotic $(\varepsilon \to 0, \gamma \to 0)$ estimates of the escape probabilities and times will be close. The sense of 'closeness' is discussed in the next section, after we introduce some concepts from the theory of large deviations.

With arbitrary WB noise models $\xi_i^\gamma(\cdot)$ such that $w_i^\gamma(\cdot)$ converges weakly to a Wiener process, the estimates of escape probabilities and times for (2.2) or (2.5) might be very different for small ε,γ. For example, if $\xi_i^\gamma(\cdot)$ is a scaled continuous parameter Markov chain or if the moments of the impulses in case 2 increase too fast [6]. One must exercise considerable care in replacing (2.2) by (2.5).

3. Large Deviations and System Approximations

Refer to (1.3) and (1.6) and let $\sigma(\cdot)$ and $f(\cdot)$ satisfy a uniform Lipschitz and a linear growth condition (i.e., $|f(x)| \leq K(1+|x|)$). Define the functionals $H(\cdot,\cdot)$ and $L(\cdot,\cdot)$ by $H(\alpha,x) = \alpha' f(x)+\alpha'\sigma(x)\sigma'(x)\alpha/2$ and $L(\beta,x) = \sup_{\alpha} [\beta'\alpha - H(\alpha,x)]$. In our PLL model the driving noise is degenerate. To prepare for this, consider the special case where $(x_1,x_2) = x$ and $\sigma_1\sigma_1'$ is uniformly positive definite and

$$(3.1) \qquad \begin{aligned} dx_1 &= f_1(x)dt + \varepsilon\sigma_1(x)dw \\ dx_2 &= f_2(x)dt \quad . \end{aligned}$$

Split $\alpha = (\alpha_1,\alpha_2)$, etc. Then

$$H(\alpha,x) = \alpha_1' f_1(x) + \alpha_2' f_2(x) + \alpha_1'\sigma_1(x)\sigma_1'(x)\alpha_1/2$$

and

$$(3.2) \qquad \begin{aligned} L(\beta,x) &= \frac{1}{2}\left(\beta_1 - f_1(x)\right)'\left(\sigma_1(x)\sigma_1'(x)\right)^{-1}\left(\beta_1 - f_1(x)\right) \\ &\qquad\qquad\qquad \text{if } \beta_2 = f_2(x) \\ &= \infty \quad \text{otherwise.} \end{aligned}$$

Define the action functional

$$(3.3) \qquad \begin{aligned} S_x(T,\phi) &= \int_0^T L\left(\dot{\phi}(s),\phi(s)\right)ds \ , \quad \text{for} \\ &\qquad\qquad \phi(0) = x \ , \quad \phi(\cdot) \text{ absolutely continuous} \\ &= \infty \quad \text{otherwise.} \end{aligned}$$

Let $x = 0$ be an asymptotically stable point of $\dot{x} = f(x)$ with G being a bounded open set whose closure $\bar{G}$ is in the domain of attraction of $\{0\}$. Define

$$(3.4)\qquad \begin{aligned} S_x^*(T) &= \inf_{\phi} S_x(T,\phi) \\ S_x^* &= \lim_{T\to\infty} S_x^*(T) \quad , \end{aligned}$$

where the inf is over all $\phi(\cdot)$ which escape G by time T. Then, under broad conditions on G [9,10]

$$(3.5a)\qquad \lim_{\varepsilon} \varepsilon^2 \log P_x\{x^\varepsilon(t) \notin G, \text{ some } t \le T\} = -S_x^*(T)$$

$$(3.5b)\qquad \lim_{\varepsilon} \varepsilon^2 \log E_x\tau_G^\varepsilon = S_0^* \ , \ x \in G \quad ,$$

If (3.5) does not hold for a particular set G, it will hold for a small perturbation of G. A maximum likelihood interpretation of (3.5) appears in the next section.

Now, let us relate the above facts to the system (2.5) Let $\theta(t) = \theta$, a constant. For notational convenience absorb the $A_0/2$ into B and the $\sigma/2$ into ε, and rewrite (2.5) as

$$(3.6)\qquad \begin{aligned} dv^\varepsilon &= Av^\varepsilon dt + B \sin(\theta-\hat{\theta}^\varepsilon)dt + B\varepsilon dw \\ (\theta\dot{-}\hat{\theta}^\varepsilon) &= -Hv^\varepsilon\, dt \quad . \end{aligned}$$

Write $x = (v,\theta-\hat{\theta})$. The set G of interest here is $G = G_1 = \{x: |\theta-\hat{\theta}| < \pi\} \cap G_2$, where G_2 is the domain of attraction of the stable point $x = 0$. The closure $\bar{G}$, is compact, but not all of it is in the domain of attraction of $\{0\}$, since $(0,\pm\pi)$ is also a singular point. Nevertheless, owing to the special form, the additional problems are minor and (3.5) holds.

The approximation (2.5) to (2.2) is valid [6] in the sense that, for $x \in G$ ($\rho \to 0$ implies $\varepsilon \to 0$, $\lambda \to 0$)

(3.7a) $$\lim_{\rho} \varepsilon^2 \log P_x \{x^\rho(t) \notin G_1 \text{ , some } t \leq T\}$$

$$= \lim_{\varepsilon} \varepsilon^2 \log P_x \{x^\varepsilon(t) \notin G_1 \text{ , some } t \leq T\}$$

(3.7b) $$\lim_{\rho} \varepsilon^2 \log E_x \tau^\rho_{G_1} = \lim_{\varepsilon} \varepsilon^2 \log E_x \tau^\varepsilon_{G_1} \quad .$$

The proof of (3.7) involves details from the theory of large deviations [6]. But without such a proof one could not justify using the simpler system (2.5) for computational purposes. The values of (1.4), (1.5) depend on appropriately timed (rare) bursts of noise and a simple diffusion approximation argument is inappropriate.

In order to relate the special form (3.1), (3.2) to (2.5), we change coordinates if needed, so that B takes the form of the column vector $(b,0,0,\cdots,0)$ and let $x_1 = v_1$, $x_2 = (v_2,\cdots,\theta-\hat{\theta})$. Then $\sigma = b = \sigma_1$.

<u>Computation of</u> $S^*_x(T)$ <u>and</u> S^*_x. Let us put the variational problem involved in (3.4) into a more enlightening form, and work with the special case (3.1), (3.2). Let $\beta_2 = f_2(x)$. Then

$$L(\beta,x) = \sup_{\alpha_1} [\alpha_1'\sigma_1(x)\sigma_1^{-1}(x)\left(\beta_1-f_1(x)\right)$$
$$- \frac{1}{2} \alpha_1'\sigma_1(x)\sigma_1'(x)\alpha_1]$$

$$= \sup_{\alpha_1} [\alpha_1'\sigma_1^{-1}(x)\left(\beta_1-f_1(x)\right) - \frac{1}{2} |\alpha_1|^2]$$

$$= |u(x)|^2/2 \quad ,$$

where $u(x) = \sigma_1^{-1}(x)\left(\beta_1-f_1(x)\right)$. Then

$$(3.8) \qquad S_x(T,\phi) = \frac{1}{2}\int_0^T |u(\phi(s))|^2\, ds \;,$$

with

$$(3.9) \qquad \dot{\phi} = f(\phi) + \sigma(\phi)u \;, \quad \phi(0) = x \;.$$

With the form (3.7), (3.8), the calculations of $S_x^*(T)$ and S_x^* are those for an optimal control problem. For our case (3.6),

$$(3.10) \qquad f(x) = \begin{bmatrix} Av + B\sin(\theta-\hat{\theta}) \\ \\ -Hv \end{bmatrix}, \; \sigma(x)u = \begin{bmatrix} bu \\ 0 \\ \vdots \\ 0 \end{bmatrix} .$$

Define $V(x) = S_x^*$. Then, purely formally, $V(\cdot)$ satisfies the Bellman equation [11,17],

$$(3.11) \qquad \min_u [L^u V(x) + u^2/2] = 0 \;, \quad x \in G_1$$

$$V(x) = 0 \;, \quad x \in \partial G_1 \;,$$

where L^u is defined by $L^u g(x) = \text{grad}_x' g(x) \cdot [f(x) + \sigma(x)u]$.

Whether or not (3.10) is formal, the <u>desired solution</u> $V(x) = S_x^*$ can be obtained via the finite difference method of [15], Chapter 9.9.6, or [16] where a method for a slightly simple problem is discussed. Let h denote a finite difference interval and e_i the unit vector in the ith coordinate direction. Define the h-Grid $G_h = G_1 \cap \{\sum_i k_i e_i h, \;\; k_i = \text{all integers}\}$.

By an appropriate choice of finite difference scheme for the derivatives in (3.10), one can obtain a finite difference solution $V^h(x)$ which converges to S_x^* uniformly in any compact set in G_1. The conditions required in [15] hold for our case (3.10), and $G = G_1$. The computational method also yields an approximation $u^h(x)$, $x \in G_h$, to the optimal control $u(x)$.

In the computation, we used the set $G_0 = \{x: |\theta-\hat{\theta}| < \pi\}$ in order not to have to calculate the domain of attraction of $x = 0$. But the result is the same, since once the path exits G_1, it will leave G_0 with $u(x) = 0$ (no additional cost).

4. Exit Paths and Measure Transformations

We work with the quantities (3.5) for the system (3.6), but there are identical interpretations for the left sides of (3.7). Let there be a finite number of minimizing functions $\{\phi_i^T(\cdot),\ i \le k\}$ for $S_x^*(T)$. Then for any $\lambda > 0$ and small ε, the overwhelming proportion of the paths which exit on the time interval $[0,T]$ starting at x, exit in $\bigcup_i N_\lambda\left(\phi_i^T(\cdot)\right)$, where N_λ is a λ-neighborhood. But the probability of exit on $[0,T]$ is small for small ε.

Let there be a finite number of minimizing functions $\bar{\phi}^i(\cdot)$, $i \le k$, for S_0^*. For our PLL problem these occur in pairs in that if $\bar{\phi}(\cdot)$ is a minimizer, so is $-\bar{\phi}(\cdot)$. (Obviously, there are an infinite number of minimizers $\bar{\phi}(\cdot)$, since we can always let $\bar{\phi}_i(\cdot) = 0$ on any interval $[0,t_1]$, $t_1 > 0$; but we do not need to count these.)

For small ε, the path $x^\varepsilon(\cdot)$ spends most time in a small neighborhood of the origin. Occasionally a large burst of noise pulls it out - but, with a very high probability, it will return to a small neighborhood of the origin before exiting from G_1. Eventually, however, the path will (w.p.1) leave G_1. When it does, loosely speaking, it leaves in a small neighborhood of the path $\{\bar{\phi}_i(\cdot)\}$. We can quantify this as follows. Let $\lambda > 0$ and let $\mu_2 > \mu_1 > 0$ be small, and let $N_{\mu_i}(0)$ denote a μ_i neighborhood of the origin. Then [6] for $x \in N_{\mu_i}(0)$,

$$(4.1)\qquad \lim_{\varepsilon} P_x\{x^\varepsilon(t) \in \bigcup_i N_\lambda(\bar{\phi}_i(t)) \text{ until the exit time } \mid x^\varepsilon(\cdot) \text{ exits } G_1 \text{ and does not return to } N_{\mu_i}(0) \text{ after leaving } N_{\mu_2}(0)\} = 1$$

There is only one pair of optimal paths. The parts of the functions

$\{\phi_i^T(\cdot),\ i\le k\}$ on any finite interval $[0,T_1]$ converge to $\{\bar{\phi}^i(\cdot),\ i\le k\}$ on $[0,T_1]$ as $T\to\infty$.

It is clear from this discussion that the most likely noise sequences which lead to exit are of the form (write $\phi(\cdot)=(\phi_1(\cdot),\phi_2(\cdot))$, via the form (3.1))

$$(4.2)\qquad w(t)\sim\frac{1}{\varepsilon}\,\sigma_1^{-1}\big(\bar{\phi}(t)\big)\Big[\bar{\phi}_1(t)-\bar{\phi}_1(0)-\int_0^t f_1\big(\bar{\phi}(s)\big)ds\Big]\quad ,$$

where $\bar{\phi}(\cdot)$ is a minimizer for S_0^*. From (4.2), we can get a clear idea of when the noise (leading to exit) will be large, and in which direction it will push the system.

<u>A likelihood function interpretation of</u> $S_x^*(T)$ <u>and</u> S_0^*. It is suggestive to view (3.4) as selecting an exit path which maximizes a likelihood function. Since for each $\varepsilon>0$, $x^\varepsilon(\cdot)$ is not differentiable, one cannot speak of a likelihood function in a strict sense. But, for the Gaussian case (3.1), one can get an intuitively reasonable maximum likelihood interpretation. Even from this the estimates (3.5) do not follow readily without the large deviations formalism.

Consider the discrete parameter process

$$(4.3)\qquad X_{i+1}^{\varepsilon,\Delta}=X_i^{\varepsilon,\Delta}+\Delta f(X_i^{\varepsilon,\Delta})+\sqrt{\Delta}\,\varepsilon\sigma(X_i^{\varepsilon,\Delta})\xi_i\quad ,$$

where $\{\xi_i\}$ are i.i.d. and normal $(0,1)$. The likelihood function of $\{X_i^{\varepsilon,\Delta},\ i\le T/\Delta\}$ evaluated on the path $X_i^{\varepsilon,\Delta}=\phi(i\Delta)$, $i\Delta\le T$, is (4.4) times a factor not depending on the path.

$$(4.4)\qquad L^\Delta(T,\phi)=\prod_0^{T/\Delta-1}\exp-\frac{\Delta}{2\varepsilon^2}\,|u(i\Delta)|^2\quad ,$$

where

$$\phi(i\Delta+\Delta)=\phi(i\Delta)+\Delta f\big(\phi(i\Delta)\big)+\Delta\sigma\big(\phi(i\Delta)\big)u(i\Delta)\quad .$$

We have

(4.5) $$S(T,\phi) = -\lim_{\Delta} \varepsilon^2 \log L^{\Delta}(T,\phi) \quad .$$

Let $x^{\varepsilon,\Delta}(\cdot)$ denote the piecewise linear interpolations of $\{X_i^{\varepsilon,\Delta}\}$ with interval Δ, and $\tau_G^{\varepsilon,\Delta}$ the escape time from G of $x^{\varepsilon,\Delta}(\cdot)$. By the standard methods of the theory of large deviations [9,10], it follows that if $\Delta \to 0$ as $\varepsilon \to 0$ and under broad conditions on G

(4.6) $$\begin{aligned} &\lim_{\varepsilon,\Delta} \varepsilon^2 \log P_x\{x^{\varepsilon,\Delta}(t) \notin G, \text{ some } t \leq T\} = -S_x^*(T) \\ &\lim_{\varepsilon,\Delta} \varepsilon^2 \log E_x \tau_G^{\varepsilon,\Delta} = -S_0^* , \quad x \in G^0 \quad . \end{aligned}$$

If (4.6) does not hold for a given G, then it will hold for a small perturbation of G. Also, (4.6) holds for the discrete time form of our special case (3.6).

These estimates provide justification for simulating the discrete parameter system in lieu of $x^{\varepsilon}(\cdot)$.

Importance sampling. Now that we know what the 'most likely exit paths' are, we are ready to define the measure transformation used to facilitate the simulations. Following the idea in [14], we transform the measure so that exit on $[0,T]$ is not a rare event. T will be large enough so that $S_0^* \sim S_0^*(T)$. To get the measure transformation, we use the approximation to the optimal controls $u^h(\cdot)$ given by the computational procedure discussed at the end of Section 3. For $x \in G_1 = G$ but not in the grid G_h, we define $u^h(\cdot)$ by a linear interpolation over the values in adjacent grid points. For simplicity, we write $u(x)$ below for $u^h(x)$.

Write the system in the form (4.7) with associated measure P_0^{ε}. We work on an arbitrary interval $[0,T]$. Define $\bar{x}^{\varepsilon}(\cdot)$ on $[0,T]$ by (4.8), with associated measure P_u^{ε}.

(4.7) $$dx^{\varepsilon} = f(x^{\varepsilon})dt + \varepsilon\sigma dw$$

(4.8) $$d\bar{x}^{\varepsilon} = f(\bar{x}^{\varepsilon})dt + \sigma u(\bar{x}^{\varepsilon})dt + \varepsilon\sigma dw \ .$$

Again, using the notation $x = (x_1, x_2)$, $\sigma = (\sigma_1, 0)$, $\cdots$, the Radon-Nikodyn derivative is [18]:

(4.9) $$\frac{dP_0^{\varepsilon}}{dP_u^{\varepsilon}}(\omega) = \exp \frac{1}{\varepsilon^2}\left[\frac{1}{2}\int_0^T |u(\bar{x}(s))|^2 ds - \int_0^T u'(\bar{x}(s))\sigma_1^{-1}[d\bar{x}_1^{\varepsilon}(s) - f_1(\bar{x}(s))ds]\right]$$

We use a discrete form in the simulations. Thus, define the processes $\{X_i^{\varepsilon,\Delta}\}$, $\{\bar{X}_i^{\varepsilon,\Delta}\}$ (with associated measures $P_0^{\varepsilon,\Delta}$, $P_u^{\varepsilon,\Delta}$, respectfully, on $[0,T/\varepsilon]$)

(4.10a) $$X_{i+1}^{\varepsilon,\Delta} = X_i^{\varepsilon,\Delta} + \Delta f(X_i^{\varepsilon,\Delta}) + \varepsilon\sqrt{\Delta}\sigma\xi_i$$

(4.10b) $$\overline{X}_{i+1}^{\varepsilon,\Delta} = \overline{X}_i^{\varepsilon,\Delta} + \Delta f(\overline{X}_i^{\varepsilon,\Delta}) + \Delta\sigma\, u(\overline{X}_i^{\varepsilon,\Delta}) + \varepsilon\sqrt{\Delta}\sigma\xi_i \ .$$

We always let $u(\overline{X}_i^{\varepsilon,\Delta}) = 0$ after the first exit time from $G = G_1$. We have, for $\psi_i^{\varepsilon,\Delta} = \overline{X}_{1,i+1}^{\varepsilon,\Delta} - \overline{X}_{1,i}^{\varepsilon,\Delta} - \Delta f_1(\overline{X}_i^{\varepsilon,\Delta})$,

(4.11) $$\frac{dP_0^{\varepsilon,\Delta}}{dP_u^{\varepsilon,\Delta}}(\omega) = \prod_0^{T/\Delta-1} \exp \frac{1}{\varepsilon^2}\left[\frac{1}{2}|u(\overline{X}_i^{\varepsilon,\Delta})|^2 \Delta - u'(\overline{X}_i^{\varepsilon,\Delta})\sigma_1^{-1}\psi_i^{\varepsilon,\Delta}\right] \ .$$

We will simulate using (4.10b). Let M denote the number of (mutually independent) runs, indexed by ω_i, $i \leq M$. Define the set A whose probability is of interest:

$$A = \{\text{paths exiting } G = G_1 \text{ by time step } T/\Delta, \text{ starting at } x(0) \approx 0\} \ .$$

Then (4.12) is an unbiased estimate of $P_0^{\varepsilon,\Delta}\{A\}$, the escape probability

from G_1 for (4.10a).

$$\frac{1}{M} \sum_{1}^{M} \frac{dP_0^{\varepsilon,\Delta}}{dP_u^{\varepsilon,\Delta}} (\omega_i) \; I_{\{\omega_i \in A\}} \quad . \tag{4.12}$$

See [14] for additional discussion on the use of such transformations, particularly for the sense in which they are optimal (in the sense that they minimize the variance of certain errors, among all such measure transformations. As seen in the following sections, the importance sampling methods works very well.

5. Simulations with $\theta(t) = \theta_0 = 0$

We use a first order filter, and the simulated PLL equations are the discrete time form of

$$
\begin{aligned}
dx_1^\varepsilon &= -ax_1^\varepsilon dt + b(\sin x_2^\varepsilon dt + \varepsilon d\omega) \quad , \\
dx_2^\varepsilon &= -x_1^\varepsilon dt \ , \quad x_2^\varepsilon = (\theta - \hat{\theta}^\varepsilon) \quad .
\end{aligned} \tag{5.1}
$$

In all cases, $a = 2$, and the results in Tables 1 to 3 are tabulated according to the value of b. In order to best compare the simulation and the theoretical estimates, we write (4.12) in the form $\exp -\overline{S}^\varepsilon(T)/\varepsilon^2$ and tabulate $\overline{S}^\varepsilon(T)$. We use $\overline{P}^\varepsilon(T)$ for the equivalent sample probability of escape on $[0,T]$. There is clearly a very close agreement between the theoretical predictions and the results of the simulations.

The (# escape) denotes the number of simulated paths escaping on $[0,T]$ under the transformed measure (for the importance sampling), not the measure of (5.1). It indicates the savings in the cost of simulation. The 'eigenvalues' denotes the eigenvalues of the noiseless (5.1), linearized at $x = 0$.

Table 1

b = 1, eigenvalues (1,1), Δ = .02, 200 Runs

T	ε	# Escape	S_0^*	$\overline{S}^{\epsilon}(T)$	$\overline{P}^{\epsilon}(T)$
3.5	.2	15	2.55	2.59	$.8 \times 10^{-28}$
	.4	83		2.55	$.95 \times 10^{-7}$
	.6	121		2.55	$.84 \times 10^{-3}$
	.8	137		2.39	$.24 \times 10^{-1}$
5	.2	42	2.55	2.56	$.16 \times 10^{-27}$
	.4	155		2.46	$.21 \times 10^{-6}$
	.6	171		2.39	$.13 \times 10^{-2}$
	.8	174		2.06	$.40 \times 10^{-1}$
	.1	184		1.80	.17
7.5	.2	129	2.55	2.5	$.81 \times 10^{-27}$
	.4	192		2.39	$.33 \times 10^{-6}$
	.6	194		2.22	$.21 \times 10^{-2}$
	.8	196		1.99	$.45 \times 10^{-1}$

Table 2

$b = 2$, eigenvalues $(-1 \pm \sqrt{i})$, $\Delta = .02$, 200 Runs

T	ε	# Escape	S_0^*	$\overline{S}^{\epsilon}(T)$	$\overline{P}^{\epsilon}(T)$
1.5	.2	5	1.28	1.36	$.18 \times 10^{-14}$
	.4	58		1.41	$.15 \times 10^{-3}$
	.6	95		1.32	$.25 \times 10^{-1}$
	.8	109		1.37	.12
2.2	.2	40	1.28	1.29	$.12 \times 10^{-13}$
	.4	128		1.25	$.42 \times 10^{-3}$
	.6	150		1.09	$.49 \times 10^{-1}$
	.8	157		.84	.23
3.3	.2	107	1.28	1.25	$.29 \times 10^{-13}$
	.4	179		1.12	$.93 \times 10^{-3}$
	.6	179		1.02	$.59 \times 10^{-1}$

Table 3

b = 0.75, eigenvalues (-1/2,-3/2), Δ = .02, 200 Runs

T	ε	# Escape	S_0^*	$\overline{S}^{\varepsilon}(T)$	$\overline{P}^{\varepsilon}(T)$
5	.2	9	3.43	3.45	$.38 \times 10^{-37}$
	.4	95		3.41	$.56 \times 10^{-9}$
	.6	124		3.39	$.82 \times 10^{-4}$
	.8	140		3.21	$.66 \times 10^{-2}$
7	.2	47	3.43	3.37	$.26 \times 10^{-36}$
	.4	156		3.35	$.82 \times 10^{-9}$
	.6	175		3.12	$.17 \times 10^{-3}$
	.8	181		2.82	$.12 \times 10^{-1}$

Comment on the Choice of T

To get $S_x^*(T)$ and the associated measure transformation, we need controls which are time dependent (hence more computation and memory is needed). In applications, the value of T is not very important (multiplying or dividing a reasonable T by a constant k also yields useful results). The ε-effects are much more pronounced. For 'large' T, $S_x^*(T) \simeq S_0^*$ and most of the 'control activity' for the optimal escape path occurs long before T. In our problem the optimal exit paths exit at $\theta-\hat{\theta} = \pm\pi$, $v = 0$, and take an infinite time to reach these points. Because of this, we chose T as follows. Using the optimal control $u(x)$ and the noiseless system, we start the trajectory very close to the origin (not at $x = 0$, since $u(x) = 0$) and let T^* denote the time the path takes to get to a position x near the exit point where S_x^* is very small. Our larger T-values are about $1.25\ T^*$.

Discussion of the Data

As can be seen from the data, even for the smaller T, the measure transformation obtained with the (nearly) optimal $u(\cdot)$ for S_x^* yields excellent results. If the prediction is close to the sample estimate for large T and small ε (say $\varepsilon = .2$), then we are very likely justified in using the small ε values of $\bar{S}^\varepsilon(T)$ as estimates of $S_0^*(T)$, and then we can compare these to the sample estimates for large values of ε.

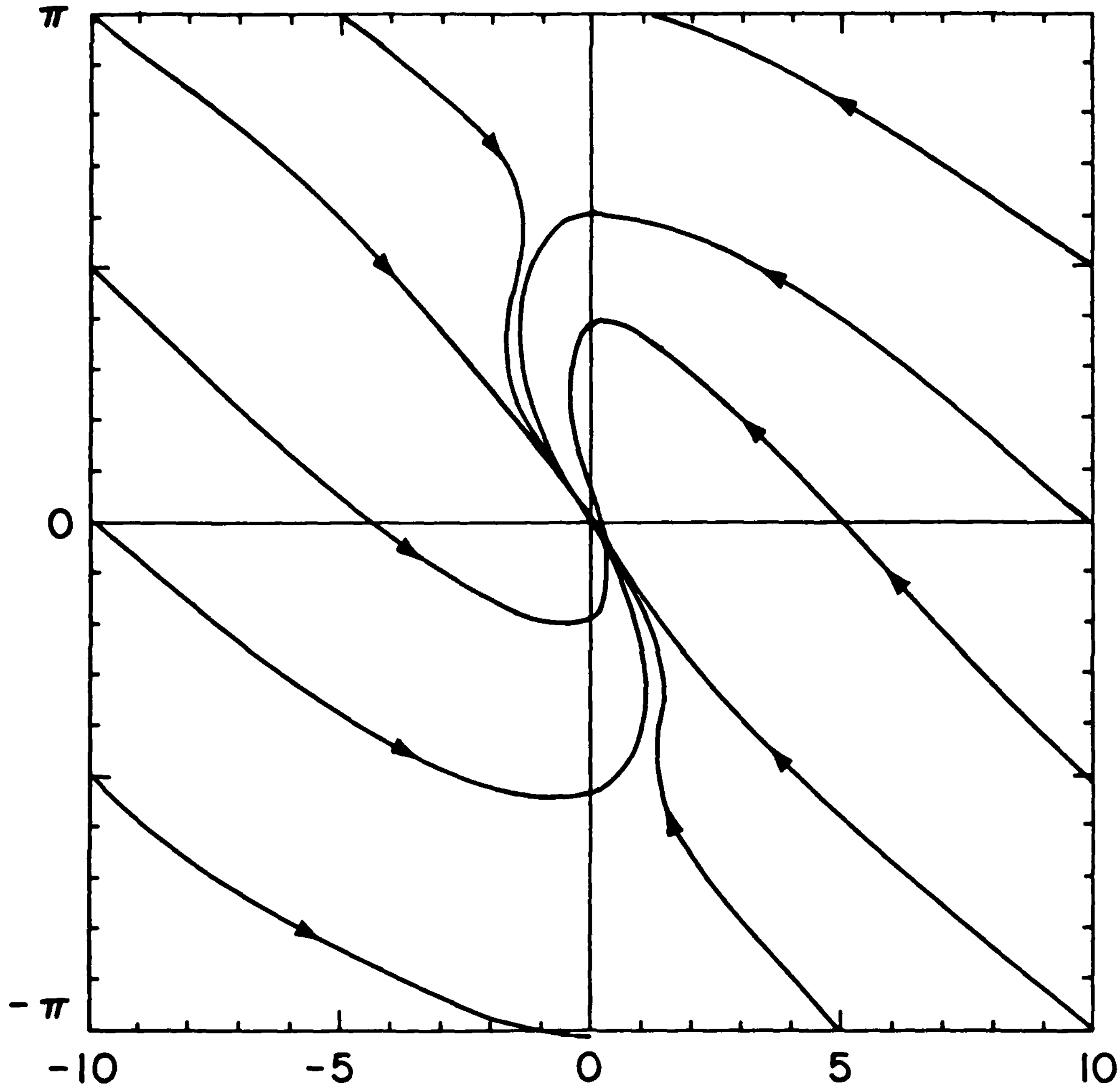
The role of the measure transformation is made very clear by the tables. E.g., for $\varepsilon = 0.2$, $b = 1$ and $T = 3.5$, the probability of escape is $\simeq 10^{-28}$. An estimate of such a quantity would require an impossible amount of computer effort. But, via the importance sampling measure trans-

formation, in 200 runs there were 15 escapes. For the transformed measure this corresponds to a (not small) probability of 0.075. Under the inverse transformation, it becomes the true estimate of the order of 10^{-28}. The method is efficient, indeed.

The size of the noise effects can be judged by the estimated escape probability. For realistic systems, one would expect small errors (perhaps of the order of 10^{-4} or smaller). Yet our predicted results are quite good even for values of ε which correspond to much larger escape probabilities.

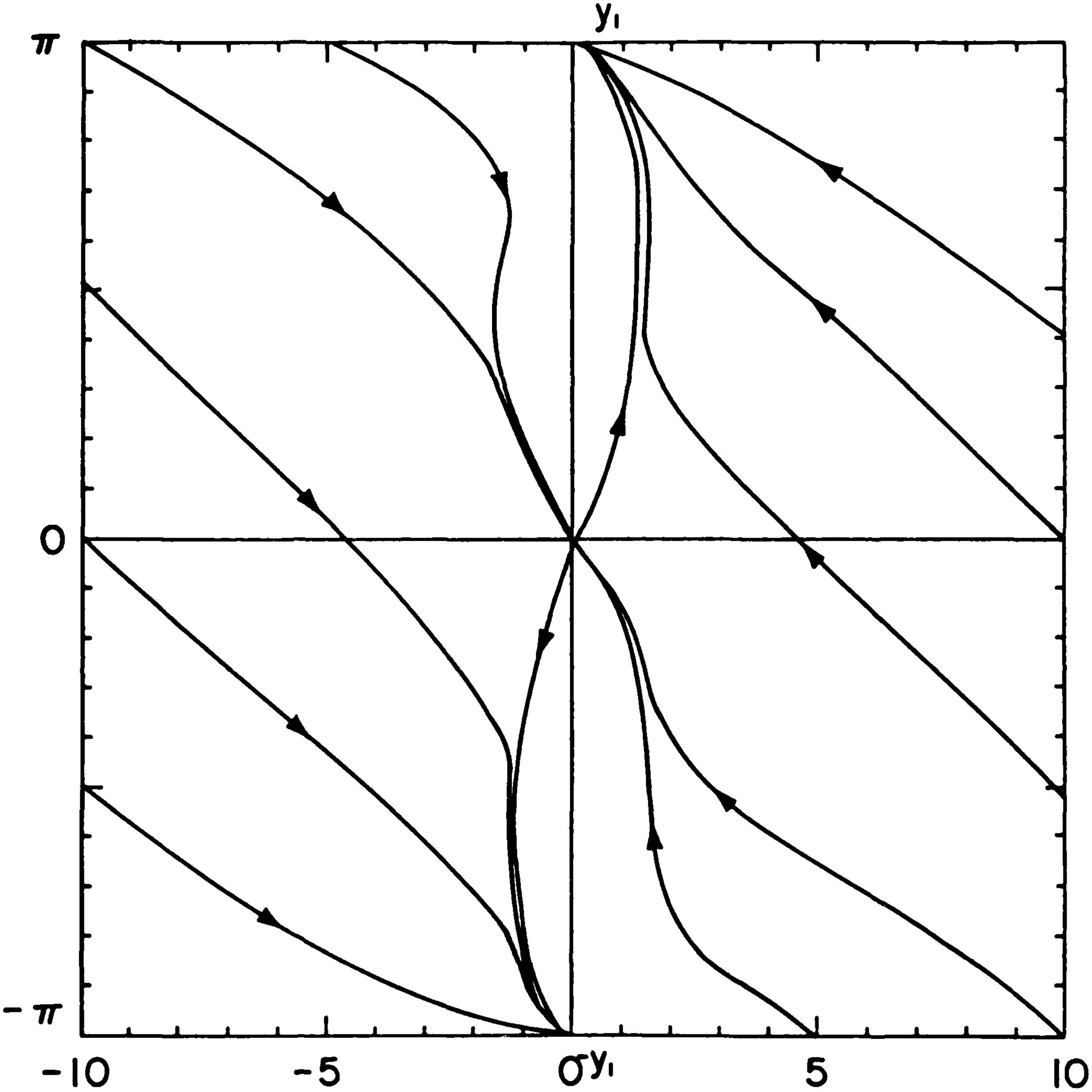
As ε increases, the theoretical predictions increasingly understate the escape probability: the noise is stronger and it is unlikely that the random escape paths of $x^{\varepsilon}(\cdot)$ are concentrated in a small tube about the optimal escape paths.

The vector field of the noiseless system (b = 1.0) is in Figure 5.1, and for the optimally controlled system in Figure 5.2. The control activity is negligible near the origin or near $x_1 = 0$, $x_2 = \pm\pi$, or for large values of x_1. The most likely noise bursts which lead to exit occur when the filter state x_1 is not large. For small ε, the exiting paths were usually in a small neighborhood of the optimal paths $(0 \to y)$, or $(0 \to -y)$, as indicated in Figure 5.2. Even when ε is rather large, the most likely exit paths are reasonably close to the optimal paths. Two typical cases are plotted in Figure 5.3, one exiting at $(0,\pi)$ and the other at $(0,-\pi)$.



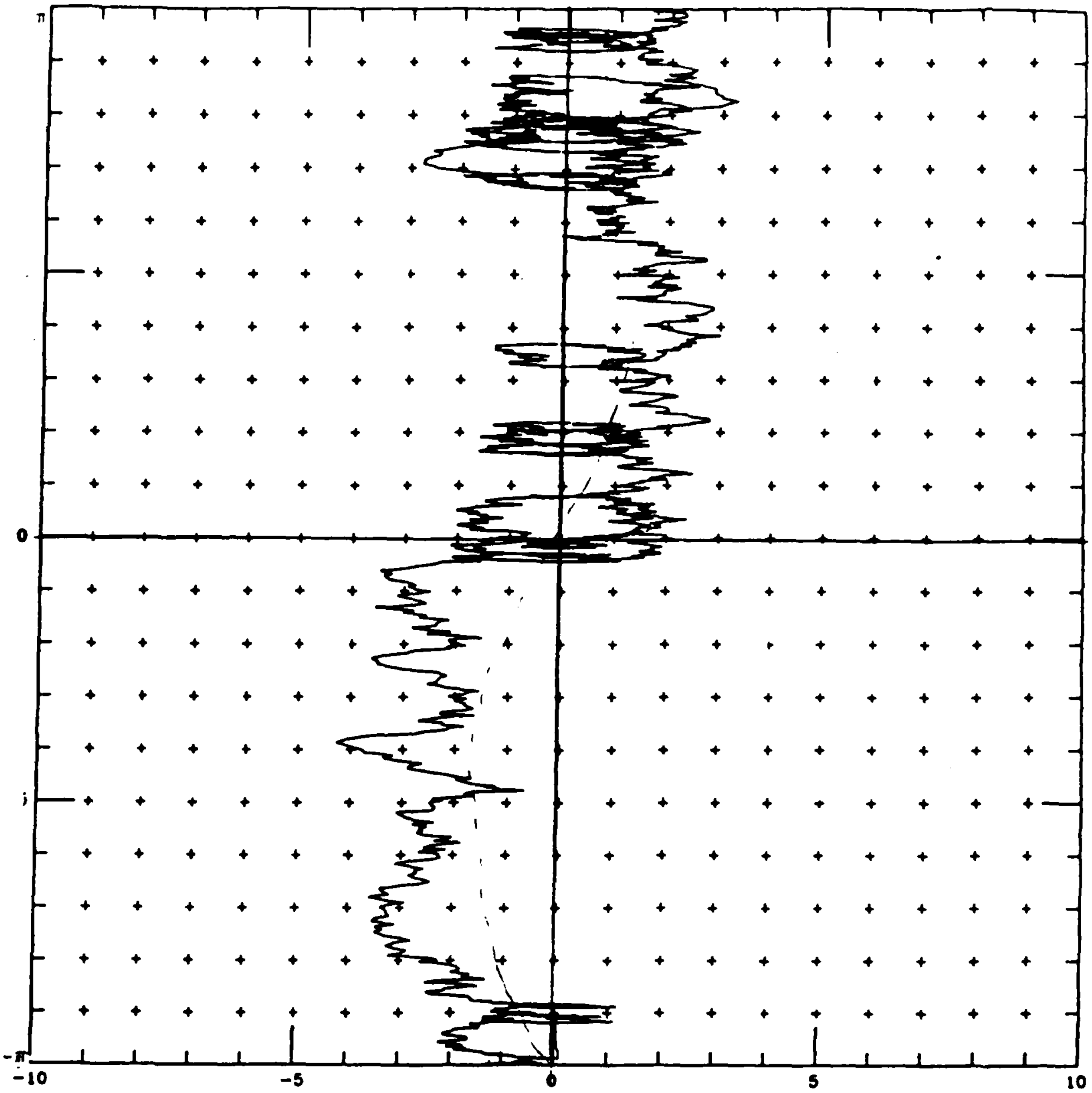
NOISELESS PATHS, b = 1.0

FIGURE 5.1



b = 1.0, OPTIMAL EXIT PATHS;
OPTIMALLY CONTROLLED PATHS

FIGURE 5.2



Two Exiting Paths

$\varepsilon = 0.6,\ \Delta = 0.005,\ b = 1.0$

Figure 5.3

6. Time Varying Phase: The Model

In order to illustrate the interactions of signals and noise which lead to the most likely escape routes, and to check the range of validity of the large deviations estimates when the signal varies with time, the following version of a pulse phase modulation system was considered. The data signal is a sequence of pulses, each of value ±1 and constant over an interval of unit length. When the sequence changes from a +1 to a -1, the transmitted signal phase $\theta(t)$ goes from $+\pi/4$ to $-\pi/4$, and conversely when the data sequence jumps from -1 to +1. The probability of changing sign is $p \in (0,1)$. The object of the PLL is to track $\theta(t)$ (hence, to track the data sequence). Often the transmitted signal for such a data sequence changes from $\pm\pi/2$ to $\mp\pi/2$, and a form of the PLL which is better suited to digital data is used. But the system used here well illustrates the general principles.

The system is (for a general filter)

$$(6.1)\qquad dv^{\varepsilon} = Av^{\varepsilon}dt + B\left(\varepsilon dw + \sin(\theta-\hat{\theta}^{\varepsilon})dt\right)$$

$$d(\theta-\hat{\theta}^{\varepsilon}) = -Hv^{\varepsilon}dt + dJ \quad ,$$

where $J(\cdot)$ is a jump process which jumps by $\pm\pi/2$ depending on the direction of change of the sign of the binary data sequence. We take the paths of $\theta(\cdot)$ and $\hat{\theta}^{\varepsilon}(\cdot)$ to be right continuous.

The comments on large deviations in Sections 3 and 4 all hold here, by proofs which are very close to those used to support the assertions in those sections. Owing to the presence of the time varying signal, the action functions <u>must depend on time as well as on the direction of the next possible change in</u> $J(\cdot)$. The appropriate action functional is

$$(6.2)\qquad S_{\tau,x}(T,\phi,J) = \frac{1}{2}\int_{\tau}^{T} |u(s)|^2 ds \quad ,$$

where

$$(6.3) \qquad d\phi = f(\phi)dt + Budt + \begin{pmatrix} 0 \\ dJ \end{pmatrix} ,$$
$$\phi(\tau) = x .$$

Here τ is the initial time, and $\phi(\tau) = x$ the inital condition and J the future jump path. $J(\cdot)$ also acts as a control force here - but is confined to be a possible signal path. Define

$$(6.4) \qquad S^*_{\tau,x}(T,\pm) = \inf_{\phi,J} S_{\tau,x}(T,\phi,J)$$

$$(6.5) \qquad S^*_{\tau,x}(\pm) = \lim_{T\to\infty} S_{\tau,x}(T,\pm) ,$$

where the inf is over all $\phi(\cdot)$ and $J(\cdot)$ satisfying (6.3) with $\phi(t) \notin G_1$ for some $t \le T$, and the <u>next possible jump</u> of $J(\cdot)$ after τ is $\pm$. Then, via mild modifications in the standard methods

$$(6.6) \qquad \lim_{\varepsilon} \varepsilon^2 \log P_{x,\tau} \{x^\varepsilon(t) \notin G_1, \text{some } \tau \le t \le T\} = -S^*_{x,\tau}(T) .$$

Also the analog of (3.7a) holds.

The limit $S^*_{\tau,x}(\pm) \equiv V(\tau,x,\pm)$ has an optimal control interpretation

$$V(\tau,x,\pm) = \min_{T,\phi,J} \{S_{x,\tau}(T,\phi,J): \phi(0) = x, \text{ (6.3) holds,}$$
$$\phi(T) \notin G_1 \text{, next possible jump after } \tau \text{ is } \pm.$$

$V(\tau,x,\pm)$ is periodic in τ; $V(n,x,\pm) = V(0,x,\pm)$, all n, and there is a possible discontinuity at τ = integer value. Because of this, $V(\tau,x,\pm)$ satisfies the Bellman equation (6.7), (6.8), for $x \in G_1$ with $V(\tau,x,\pm) = 0$ for $x \notin G_1$.

$$(6.7) \qquad 0 = \min_u \left[\frac{\partial V(\tau,x,\pm)}{\partial \tau} + L^u V(\tau,x,\pm) + u^2/2\right] \quad \text{for } 0 \le \tau < 1 ,$$

$$V(1^-,x,\pm) = \min \{V(0,x\pm\binom{0}{\pi/2},\mp),\ V(0,x,\pm)\} \tag{6.8}$$

The numerical procedure alluded to in Section 3 can still be used-- We use the state variable triple $(\tau,x,\pm)$, with $x \in G_1$ and $\tau \in [0,T]$. All the comments in Section 4 concerning the most likely exit paths continue to hold, except, of course the exit paths and solutions to the variational problems might have jumps at integer values of time.

The importance sampling method and measure transformation are also almost identical to those described in Section 4. We work with the discrete parameter systems on $[0,T]$ (with measures $P_0^{\varepsilon,\Delta}$ and $P_u^{\varepsilon,\Delta}$, respectively, where u is the optimal policy obtained from (6.7), (6.8)).

$$\bar{X}_{i+1}^{\varepsilon,\Delta} = \bar{X}_i^{\varepsilon,\Delta} + \Delta f(\bar{X}_i^{\varepsilon,\Delta}) + \begin{pmatrix} b \\ 0 \end{pmatrix} u + \varepsilon\sqrt{\Delta} \begin{pmatrix} b \\ 0 \end{pmatrix} \xi_i + \begin{pmatrix} 0 \\ J(i\Delta+\Delta)-J(i\Delta) \end{pmatrix} \tag{6.9}$$

$$X_{i+1}^{\varepsilon,\Delta} = X_i^{\varepsilon,\Delta} + \Delta f(X_i^{\varepsilon,\Delta}) + \varepsilon\sqrt{\Delta} \begin{pmatrix} b \\ 0 \end{pmatrix} \xi_i + \begin{pmatrix} 0 \\ J(i\Delta+\Delta)-J(i\Delta) \end{pmatrix} \tag{6.10}$$

Then with $u(x,t,j)$ denoting the optimal control at time t, when the next possible jump can be j, we use the measure transformation

$$\frac{dP_0^{\varepsilon,\Delta}}{dP_u^{\varepsilon,\Delta}}(\omega) = \prod_0^{T/\Delta-1} \exp \frac{1}{\varepsilon^2}\Big(\frac{1}{2}\,|u(\bar{X}_i^{\varepsilon,\Delta},i\Delta,j(i\Delta))|^2\Delta - u'(\bar{X}_i^{\varepsilon,\Delta},i\Delta,j(i\Delta))\sigma_1^{-1}\psi_i^{\varepsilon,\Delta}\Big)\ , \tag{6.11}$$

where $\sigma_1 = b$ and $\psi_i^{\varepsilon,\Delta} = \bar{X}_{1,i+1}^{\varepsilon,\Delta} - \bar{X}_{1,i}^{\varepsilon,\Delta} - \Delta f_1(\bar{X}_i^{\varepsilon,\Delta})$.

T^* was selected as in Section 5 for the constant phase system.

7. Simulations With Pulsed Phase

The simulations all started with $x(0)$ at zero, and with the next possible jump in phase being positive. Two types of jump sequences $J(\cdot)$ were used. In the first, denoted by 'random jumps,' the phase changed by $\pm\pi/2$ at each unit of time with probability $p = 1/2$. Solving the variational problem for $S^*_{0,0}(+)$ yields a unique minimizing pair $(\bar{\phi}(\cdot),\bar{J}(\cdot))$, where $\bar{\phi}(\cdot)$ is an 'optimal escape path', starting at $x = 0$ at time $\tau = 0$. The $\bar{J}(\cdot)$ might have one or many jumps in it. When $\bar{J}(\cdot)$ was used in lieu of random jumps, the results are labeled 'optimal jumps' in the tables below.

Tables 4 to 6 list the results of the simulations $\{x(0) = 0,\ \tau(0) = 0\}$, and excellent agreement between the sample estimates and the theoretical predictions. For example, in Table 5, where $\varepsilon = .6$ and the estimated escape probability is the very high 0.37, there is still good agreement.

The results of the simulations with the 'optimal jumps' should be somewhat more consistent with the theoretical values of $S^*_{0,0}(\pm)$ than would the 'random jump' simulations, since it is the optimal jumps which are used to compute $S^*_{0,0}(\pm)$. But agreement is good for both cases. The sample $\bar{S}^{\varepsilon}(T)$ values for the random jump case are larger than for the optimal jump case, as expected, since the optimal jumps are calculated to help force the system out of G_1. But, as $\varepsilon \to 0$, the two estimates are close. For small ε, the exit paths were in a small neighborhood of the optimal exit path. Two optimal exit paths are plotted in Fig. 7.1, for $b = 1$ and $b = 0.75$, when the next possible jump is $+\pi/2$ in the first case, and $-\pi/2$ in the second case.

For the case $b = 1$, the optimal path waits until time $\tau = 1$, with zero control $u(x)$ then jumps to $(0,\pi/2)$, then a large control effort moves it close to the optimal exit point $(0,\pi)$, during which time there are no further jumps. For the

case $b = 0.75$, the optimal exit path involves waiting at the origin with zero control until a time $T_1 < 1$, then being controlled to y_1 in $[T_2,1]$, then jumping to y_2 at $\tau = 1$ and finally having a burst of control effort which moves it close to the optimal exit point y_3, during which there are no further jumps. The intervals of heavy control effort (and the corresponding directions) correspond to the most likely intervals during which a burst of noise (and its corresponding direction) will lead to exit. The first burst of noise on $[T_2,1]$ positions the path to take best advantage of the next jump. Of course, even if the first burst of noise occurs 'on schedule', the next jump might not occur 'on schedule'. Then, most likely, the system will drift back to near the origin. What we have described are the paths in small neighborhoods of which exit is most likely to occur. The sample paths some time before the 'final sequence' leading to exit might be rather complex--owing to the jumps.

Table 4

Eigenvalue (-1,-1), Δ = 0.025, b = 1

T^* = 10 Sec, 500 Runs

Optimal Jumps:

T	ε	# Escape	$S^*_{0,0}(\pm)$	$\overline{S}^\epsilon(T)$	$\overline{P}^\epsilon(T)$
10	.2	24	.79	.82	$.13 \times 10^{-8}$
	.3	145		.82	$.11 \times 10^{-3}$
	.4	251		.79	$.72 \times 10^{-2}$
	.5	335		.71	$.58 \times 10^{-1}$
	.6	399		.63	.17

Random Jumps, p = 1/2

10	.2	25	.79	.83	$.11 \times 10^{-8}$
	.3	125		.87	$.66 \times 10^{-4}$
	.4	219		.86	$.48 \times 10^{-2}$
	.5	265		.86	$.32 \times 10^{-1}$
	.6	312		.84	$.97 \times 10^{-1}$
15	.2	64	.79	.81	$.17 \times 10^{-8}$
	.4	274		.82	$.59 \times 10^{-2}$
5	.2	10	.79	.84	$.82 \times 10^{-9}$
	.4	95		.98	$.22 \times 10^{-2}$

Table 5

(2,2) Eigenvalues $(-1 \pm \sqrt{i})$, $\Delta = 0.025$, $b = 2$

$T^* = 6$ Sec, 500 Runs

Random Jumps, $p = 1/2$

T	ε	# Escape	$S^*_{0,0}(\pm)$	$\bar{S}^{\varepsilon}(T)$	$\bar{P}^{\varepsilon}(T)$
6	.2	86	.40	.40	$.5 \times 10^{-4}$
	.3	187		.41	$.11 \times 10^{-1}$
	.4	233		.44	$.63 \times 10^{-1}$
	.5	299		.41	.20
	.6	367		.36	.37
	.7	407		.31	.53
8	.3	228	.40	.38	$.14 \times 10^{-1}$

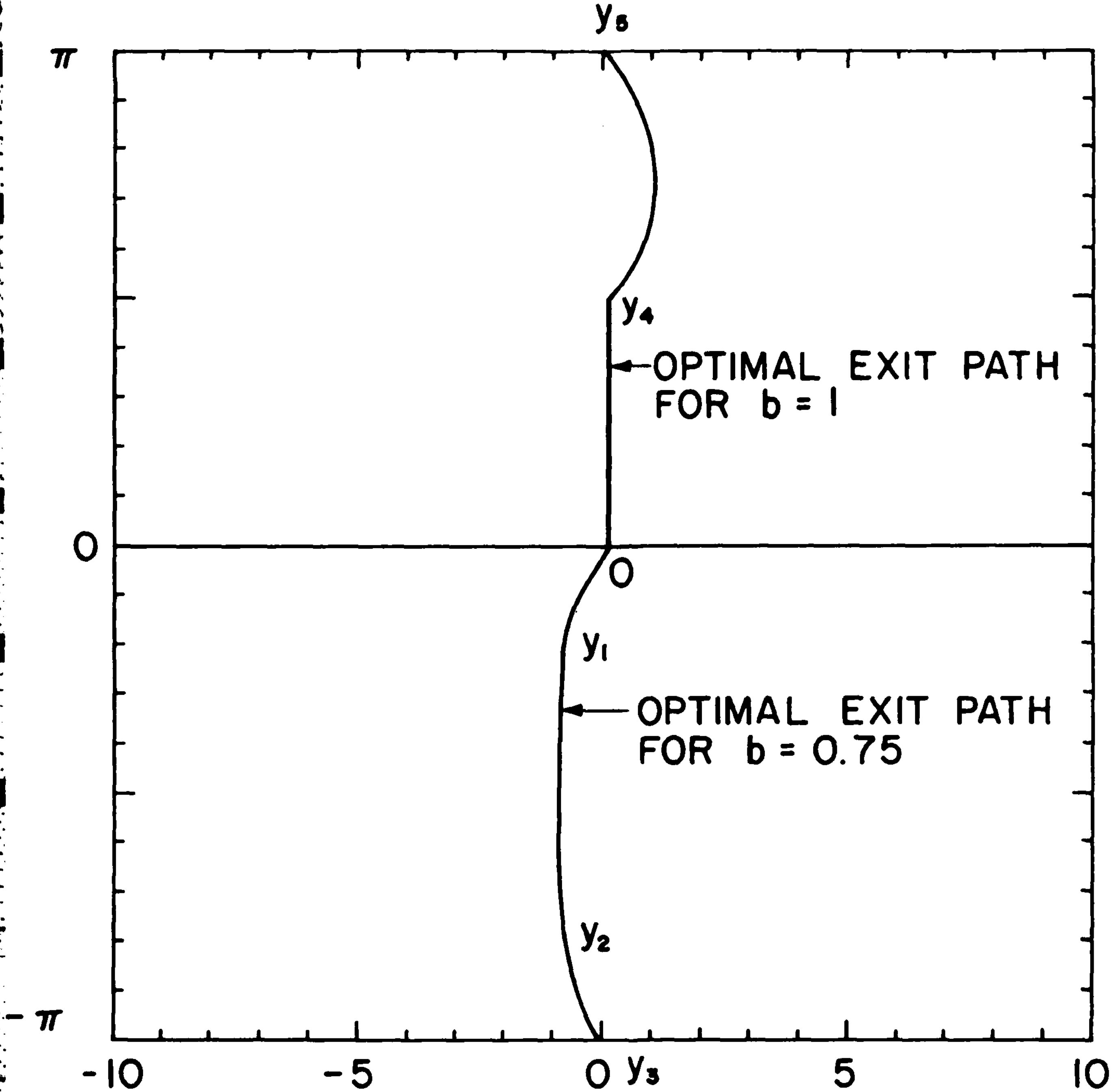
Table 6

Eigenvalues (-1/2,-3/2), $\Delta = 0.025$, $b = 0.75$

$T^* = 10$ Sec, 500 Runs

Random Jumps, $p = 1/2$

T	ε	# Escape	$S^*_{0,0}(\pm)$	$\overline{S}^\varepsilon(T)$	$\overline{P}^\varepsilon(T)$
10	.2	180	1.15	1.11	$.84 \times 10^{-12}$
	.3	235		1.16	$.26 \times 10^{-5}$
	.4	279		1.24	$.42 \times 10^{-3}$
	.5	315		1.18	$.88 \times 10^{-2}$
	.6	319		1.23	$.33 \times 10^{-1}$
	.7	354		1.19	$.88 \times 10^{-1}$
	.8	373		1.14	.17
	.9	409		1.08	.26
8	.2	124	1.15	1.11	$.94 \times 10^{-12}$
	.3	183		1.20	$.17 \times 10^{-5}$
	.4	242		1.24	$.44 \times 10^{-3}$
	.5	239		1.28	$.61 \times 10^{-2}$



OPTIMAL EXIT PATHS

FIGURE 7.1

Figure 7.1 indicates an optimal exit path, but it does provide a picture of where and when the controls are strongest (equivalently, where and when the noise bursts which lead to exit are likely to occur). We will now describe part of the time dependent control behavior, for the case $b = 1$ on the interval $[0,1)$, when the next possible jump in phase $(\theta-\hat{\theta})$ must be negative. Since the control is added to the uncontrolled dynamics, it will be useful to refer to Fig. 5.1.

There are 2 regions where the effects of the control are large and its effects obvious. In a small neighborhood below and to the right of $(0,\pi)$ (in the phase plane) the controls are large and positive and essentially independent of time. Here, the control is trying to complete the job of pushing the path out of G_1. The relatively small time dependence is probably due to the fact that for the deterministic optimal control problem, there need not be another jump. If the process is close to $(0,\pi)$, it is 'cheaper' to treat it as though there were no additional jumps. The second area where the controls are large is in the brood region below the origin, but here the behavior is quite time dependent. The controls act to counteract the unforced system's damping, so as to position the process such that a jump of $-\pi/2$ will (or nearly will) drive the path out of G_1. Directly below the origin the control takes negative values, which increase as $\tau \to 1$. The dominant effect here is to 'slow down' the path and prevent it from being pushed into the right half plane, where it would be swept upward and further away from the $-\pi/2$ phase level before the next 'negative' jump occurs.

References

[1] A.J. Viterbi, Principles of Coherent Communication, 1966, McGraw-Hill, New York.

[2] W.C. Lindsey, and M.K. Simon, Telecommunication Systems Engineering, 1973, Prentice-Hall, Englewood Cliffs, N.J.

[3] F.M. Gardner, Phaselock Techniques, 1979, Wiley, New York.

[4] W.C. Lindsey, Synchronization Systems in Communication, 1972, Prentice-Hall, Englewood Cliffs, New Jersey.

[5] H. Kushner, "A cautionary note on the use of singular perturbations", Stochastics, 6, 1982, 117-120.

[6] P. Dupuis, H.J. Kushner, "Large deviations estimates for systems with small noise effects, and applications to stochastic systems theory", to appear SIAM J. on Control and Optimization; Brown Univ., LCDS Rep. #84-34

[7] D. Ludwig, "Persistence of dynamical systems under random perturbations", SIAM Rev, 17, (1975), 605-610.

[8] B. Matkowsky, and Z. Schuss, "The exit problem for randomly perturbed dynamical systems", SIAM J. on Appl. Math, 33, 1977, 365-382.

[9] S.P.S. Varadhan, Large Deviations and Applications, CBMS-NSF Regional Conference Seminar, SIAM, Philadelphia, 1984.

[10] M.I. Freidlin, and A.D. Ventsell, Random Perturbations of Dynamical Systems, Springer, Berlin, 1984.

[11] M.I. Freidlin, "The averaging principle and theorems on large deviations", Russian Math. Surveys, 33, July-Dec., 1978, pp. 117-176.

[12] W.H. Fleming, "Stochastic control for small noise intensities", SIAM J. Control, 9, 1971, 473-517.

[13] B.Z. Bobrovsky, and Z. Schuss, "A singular perturbation method for the computation of the mean first passage time in a non-linear filter", SIAM J. on Appl. Math., 42, 1982, p. 174-187.

[14] M. Cottrell, J.C. Fort, and G. Malgouynes, "Large deviations and rare events in the study of stochastic algorithms", IEEE Trans. on Automatic Control, AC-28, 1983, 907-920.

[15] H.J. Kushner, Probability Methods for Approximations in Stochastic Control and for Elliptic Equations, 1977, Academic Press, New York.

[16] H.J. Kushner, "A survey of some applications of probability and stochastic control theory to finite difference methods for degenerate elliptic and parabolic equations", SIAM Rev., Oct, 1976, 18, 545-577.

[17] W.H. Fleming, and R.W. Rishel, Deterministic and Stochastic Optimal Control, 1975, Springer, Berlin.

[18] R.S. Lipster, and A.N. Shiryaev, Statistics of Random Processes, 1977, Springer, Berlin.

[19] H.J. Kushner, Approximation and Weak Convergence Methods for Random Processer with Applications to Stochastic Systems Theory, M.I.T. Press, Cambridge, U.S.A.

[20] G. Blankenship, and G.C. Papanicolaou, "Stability and control of stochastic systems with wide band noise disturbances", SIAM J. Appl. Math., 34, 1978, 437-476.

END
FILMED

5-86

DTIC